Deep Reinforcement Learning with Applications in Transportation

Zhiwei (Tony) Qin
DiDi AI Labs, Didi Chuxing
Mountain View, California
qinzhiwei@didiglobal.com

Jian Tang
DiDi AI Labs, Didi Chuxing
Beijing, China
Syracuse University
Syracuse, New York
tangjian@didiglobal.com

Jieping Ye
DiDi AI Labs, Didi Chuxing
Beijing, China
University of Michigan
Ann Arbor, Michigan
yejieping@didiglobal.com

ABSTRACT

This tutorial aims to provide the audience with a guided introduction to deep reinforcement learning (DRL) with specially curated application case studies in transportation. The tutorial covers both theory and practice, with more emphasis on the practical aspects of DRL that are pertinent to tackle transportation challenges. Some core examples include online ride order dispatching, fleet management, traffic signals control, route planning, and autonomous driving.

CCS CONCEPTS

• **Computing methodologies → Sequential decision making**; **Multi-agent reinforcement learning**; **Markov decision processes**; *Neural networks.*

KEYWORDS

deep reinforcement learning, intelligent transportation, ridesharing

ACM Reference Format:
Zhiwei (Tony) Qin, Jian Tang, and Jieping Ye. 2019. Deep Reinforcement Learning with Applications in Transportation. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3292500.3332299

1 INTRODUCTION

Transportation, particularly the mobile ride-sharing domain has a number of traditionally challenging dynamic decision problems that have long threads of research literature and readily stand to benefit tremendously from artificial intelligence (AI). Some core examples include online ride order dispatching, which matches available drivers to trip requesting passengers on a ride-sharing platform in real-time; route planning, which plans the best route between the origin and destination of a trip; and traffic signals control, which dynamically and adaptively adjusts the traffic signals within a region to achieve low delays. All of these problems have a common characteristic that a sequence of decisions is to be made while we care about some cumulative objectives over a certain horizon. Reinforcement learning (RL) is a machine learning paradigm that trains an agent to learn to take optimal actions (as measured by the total cumulative reward achieved) in an environment through interactions with it and getting feedback signals. It is thus a class of optimization methods for solving sequential decision-making problems. Thanks to the rapid advancement in deep learning research and computing capabilities, the integration of deep neural networks and RL has generated explosive progress in the latter for solving complex large-scale learning problems, attracting huge amount of renewed interests in the recent years. The combination of deep learning and RL has even been considered as a path to true AI. It presents a tremendous potential to solve some hard problems in transportation in an unprecedented way.

KDD '19, August 4–8, 2019, Anchorage, AK, USA

ACM ISBN 978-1-4503-6201-6/19/08.
https://doi.org/10.1145/3292500.3332299

2 TARGET AUDIENCE AND PREREQUISITES

This tutorial is targeted to researchers and practitioners with a general machine learning background and are interested in working on applications of deep RL (DRL) in transportation. The goal of this tutorial is to provide the audience with a guided introduction to this exciting area of AI with specially curated application case studies in transportation. The tutorial covers both theory and practice, with more emphasis on the practical aspects of DRL that are pertinent to tackle transportation challenges. After the half-day of lectures, the audience would get an overview of the core DRL methods and their applications, particularly in transportation and ride-sharing domains. They will have a better understanding about the major challenges in transportation and how DRL can help solve those problems. They will also be introduced to several popular open-source DRL development and benchmarking frameworks to get a head-start in experimentation. The prerequisite knowledge assumed of the audience includes basic understanding of deep neural networks, optimization, and machine learning concepts. Exposure to Markov decision process and operations research in general is preferred.

3 TUTORIAL OUTLINE

We will deliver the tutorial in three parts. Part 1 starts from the basic elements of RL and tabular learning algorithms to prepare the audience with a good foundation. Function approximation is explained next to lead to the value function based DRL methods, such as the deep Q-network. In Part 2 of the tutorial, we will cover the class of RL methods that learn an optimal policy directly, namely the actor-critic and policy gradient methods. Additionally, we will discuss advanced topics like transfer learning in RL and multi-agent RL. Throughout both Parts 1 and 2, we will deep dive into application examples in transportation based on real-world data published by Didi Chuxing, the world?s leading mobile transportation platform. The final Part 3 will be devoted to getting the audience introduced to several popular open-source frameworks for developing and benchmarking RL algorithms. The tutorial will close with brief discussions on a few other industrial applications. We will also offer useful and complementary information to the SIGKDD community for those interested in pursuing this research area.

3.1 Basics and value-based methods

Machine learning paradigms: supervised, unsupervised, RL
RL basics

- Markov decision process
- Optimization problem, objective
- Value function, policy
- DP methods: value iterations, policy iterations
- TD learning [14]
- Q-learning [21], SARSA [15]
- *Example: TD(0) policy improvement for order dispatching* [22]

Function approximation [18]

- Linear/ Non-linear approximation, neural networks
- DQN [10], double DQN [19], Deep SARSA [3, 24]
- Experience replay: prioritized experience replay [12]
- *Example: Deep value networks for dispatching* [17, 20]
- *Example: DQN for dispatching and repositioning* [4]
- *Example: DQN for driver repositioning $fleetmanagement$*
- *Example: DQN for carpool decision-marking* [5]

3.2 Policy-based methods and advanced topics

Policy optimization

- Policy gradient: REINFORCE, DDPG [8], PPO [13]
- Actor-critic: A2C, A3C [9]
- *Example: Autonomous driving control*
- *Example: Route planning/navigation with and without maps*

Advanced topics

- Transfer learning, *Examples: Transfer among cities for dispatching* [20]; City navigations
- Multi-agent RL: mean-field [23], *Example: mean-field MARL for dispatching* [6]

3.3 Practice

- RL development frameworks
- Application specific tools
  - Traffic Lights Control: SUMO, Flow
  - Autonomous Driving: TORCS, CARLA
- Open data sets

In addition to the papers referenced above, the following textbooks [1, 2, 11, 15, 16] and survey paper [7] form the general basic references for this tutorial.

4 ACKNOWLEDGMENTS

The authors would like to thank Lulu Zhang from DiDi Research Outreach for the managerial contribution to this tutorial program.

REFERENCES

[1] Dimitri P Bertsekas, Dimitri P Bertsekas, Dimitri P Bertsekas, and Dimitri P Bertsekas. 2005. *Dynamic programming and optimal control*. Vol. 1. Athena scientific Belmont, MA.
[2] Dimitri P Bertsekas and John N Tsitsiklis. 1995. Neuro-dynamic programming: an overview. In *Proceedings of the 34th IEEE Conference on Decision and Control*, Vol. 1. IEEE Publ. Piscataway, NJ, 560–564.
[3] Michael Ganger, Ethan Duryea, and Wei Hu. 2016. Double Sarsa and Double Expected Sarsa with Shallow and Deep Learning. *Journal of Data Analysis and Information Processing* 4, 04 (2016), 159.
[4] J. Holler, Z. Qin, X. Tang, Y. Jiao, T. Jin, S. Singh, C. Wang, and J. Ye. 2018. Deep Q-Learning Approaches to Dynamic Multi-Driver Dispatching and Repositioning. In *NeurIPS 2018 Deep Reinforcement Learning Workshop*.
[5] Ishan Jindal, Zhiwei Tony Qin, Xuewen Chen, Matthew Nokleby, and Jieping Ye. 2018. Optimizing Taxi Carpool Policies via Reinforcement Learning and Spatio-Temporal Mining. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 1417–1426.
[6] Minne Li, Zhiwei Qin, Yan Jiao, Yaodong Yang, Zhichen Gong, Jun Wang, Chenxi Wang, Guobin Wu, and Jieping Ye. 2019. Efficient Ridesharing Order Dispatching with Mean Field Multi-Agent Reinforcement Learning. In *To appear in Proceedings of the 2019 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee.
[7] Yuxi Li. 2017. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274* (2017).
[8] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971* (2015).
[9] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*. 1928–1937.
[10] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.
[11] Warren B Powell. 2007. *Approximate Dynamic Programming: Solving the curses of dimensionality*. Vol. 703. John Wiley & Sons.
[12] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. 2015. Prioritized experience replay. *arXiv preprint arXiv:1511.05952* (2015).
[13] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).
[14] Richard S Sutton. 1988. Learning to predict by the methods of temporal differences. *Machine learning* 3, 1 (1988), 9–44.
[15] Richard S Sutton, Andrew G Barto, et al. 1998. *Reinforcement learning: An introduction*. MIT press.
[16] Csaba Szepesvári. 2010. Algorithms for reinforcement learning. *Synthesis lectures on artificial intelligence and machine learning* 4, 1 (2010), 1–103.
[17] Xiaocheng Tang, Zhiwei (Tony) Qin, Fan Zhang, Zhaodong Wang, Zhe Xu, Yintai Ma, Hongtu Zhu, and Jieping Ye. 2019. A Deep Value-network Based Approach for Multi-Driver Order Dispatching. In *To appear in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
[18] John N Tsitsiklis and Benjamin Van Roy. 1997. Analysis of temporal-diffference learning with function approximation. In *Advances in neural information processing systems*. 1075–1081.
[19] Hado Van Hasselt, Arthur Guez, and David Silver. 2016. Deep Reinforcement Learning with Double Q-Learning.. In *AAAI*. 2094–2100.
[20] Zhaodong Wang, Zhiwei Qin, Xiaocheng Tang, Jieping Ye, and Hongtu Zhu. 2018. Deep Reinforcement Learning with Knowledge Transfer for Online Rides Order Dispatching. In *International Conference on Data Mining*. IEEE.
[21] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8, 3-4 (1992), 279–292.
[22] Zhe Xu, Zhixin Li, Qingwen Guan, Dingshui Zhang, Qiang Li, Junxiao Nan, Chunyang Liu, Wei Bian, and Jieping Ye. 2018. Large-Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 905–913.
[23] Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. 2018. Mean Field Multi-Agent Reinforcement Learning. *CoRR* abs/1802.05438 (2018). arXiv:1802.05438 http://arxiv.org/abs/1802.05438
[24] Dongbin Zhao, Haitao Wang, Kun Shao, and Yuanheng Zhu. 2016. Deep reinforcement learning with experience replay based on SARSA. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*. IEEE, 1–6.